

## Modelling the feedback effects of reconfiguring health services

### Abstract

The shift in the balance of health care, bringing services ‘closer to home’, is a well established trend, which has been motivated by the desire to improve the provision of services. However, these efforts may be undermined by the improvements in access stimulating demand. Existing analyses of this trend have been limited to isolated parts of the system with calls to control demand with stricter clinical guidelines or to meet demand with capacity increases. By failing to appreciate the underlying feedback mechanisms, these interventions may only have a limited effect. We demonstrate the contribution offered by system dynamics modelling by presenting a study of two cases of the shift in cardiac catheterisation services in the U.K. We hypothesise the effects of the shifts in services and produce model output that is not inconsistent with real world data. Our model encompasses several mechanisms by which demand is stimulated. We use the model to clarify the roles for stricter clinical guidelines and capacity increases, and to demonstrate the potential benefits of changing the goals that drive activity.

Key words: health service; planning; simulation; system dynamics

## Introduction

The balance of health care is shifting towards the primary level, bringing services ‘closer to home’<sup>1,2</sup>. This trend has been motivated by the broad desire to improve the provision of services. More specifically, there is the wish to expand services. By relocating services and reallocating responsibilities, service shifts enable traditional capacity restrictions to be overcome. There is also the desire to improve the quality of services. For example, the associated capacity increases can provide reduced waiting times. Service shifts can also lead to less fragmented care by consolidating services within fewer service levels and sometimes even within a single service centre.

However, the efforts to improve services overall may be undermined by the unintended consequences of service shifts, in particular, demand being stimulated by the improvements in access to services. By access influencing demand as well as the reverse, a feedback structure (or closed causal relationship) is formed. This general feedback phenomenon has been frequently reported in the health care literature (e.g.<sup>3-8</sup>). Demand increases are particularly problematic in the U.K. health care system, the National Health Service (NHS), where there are persistent concerns about excessive waiting lists and waiting times, and rising costs. Further complexity arises with the varying views about the value of stimulated demand<sup>8-10</sup>.

The evidence of shifts in services influencing demand has tended to be anecdotal or unsubstantiated assertions with existing analyses limited to isolated parts of the system, based upon economic appraisals and surveys. Insufficient attention has been given to the underlying feedback mechanisms. There have also been suggestions, but little formal analysis, of further ways in which demand may be stimulated. For example, Hamblin *et al*<sup>11</sup> describe how when staff gain new skills, as a result of taking on new duties, they identify more patients in need of treatment. It is not clear how the various feedback mechanisms work. Furthermore, existing analyses have led to calls to control demand by introducing stricter clinical guidelines or to meet demand with capacity increases<sup>12-14</sup>. However, these interventions may only have a limited impact due to the complexity of the underlying feedback mechanisms. By understanding these better, it may be possible to design more effective policy interventions.

Exploring the effects of policy changes and experimenting with alternative policy formulations is not feasible in the real world. However, simulation models can provide a flexible experimental environment for this purpose. The system dynamics method<sup>15</sup> is specifically designed for the analysis of feedback mechanisms. This emphasis on feedback produces quite a different simulation approach compared with the more common discrete event simulation<sup>16</sup> (See Jun *et al*<sup>17</sup> for examples of the latter). For example, in system dynamics, in order to discern the feedback mechanisms, it is necessary to have an appropriate conceptual distance (i.e. to ‘stand back’ from the detail). Therefore, system dynamics models tend to be more holistic, with aggregate flows rather than focusing on the complexities of flows of individuals as in discrete event simulation. Furthermore, the aim of system

dynamics is to provide an understanding of the modes of behaviour (the basic trends). The aims of discrete event simulation are more quantitative with their models repeatedly re-run to provide predictions. In this paper, we describe a system dynamics modelling study<sup>18</sup> of two cases of a specific shift across the patient referral chain in the NHS.

The patient referral chain runs from the primary level (care by general practitioners or GPs) to the secondary level (district general hospitals) and from there on to the tertiary level (more specialised hospital-based care carried out at specialist centres). Our study considered the shift in cardiac catheterisation (CC) services from the tertiary level to the secondary level for low risk investigations. A CC investigation is an invasive procedure used to diagnose heart disease. The majority of elective (non-emergency) investigations are low risk. Whilst a few patients who present at hospital as emergencies will subsequently stabilise and become elective referrals for a CC investigation, elective referrals typically arise from patients first seen at a hospital outpatient appointment. A CC investigation is not carried out as part of the outpatient appointment but at a later date, the delay being determined by the urgency of the investigation and the patient's health authority's funding constraints. We considered the development of CC services at two English district general hospitals. For reasons of confidentiality, they are referred to as 'Veinbridge Hospital' and 'Ribsley Hospital'.

The model conceptualisation process involved collecting archival data, observational work, informal discussions with junior hospital staff, and formal interviews with senior health professionals including consultant cardiologists, hospital managers and health service purchasers. The purchasers were senior staff who worked at the Ribsley and Veinbridge health authorities and were thus responsible for commissioning health services for the populations in these respective districts. The model's target audience was primarily strategic decision makers involved in development of CC services at Ribsley and Veinbridge, and thus was tailored to address the concerns of these senior health professionals.

In the remaining sections of this paper, we first briefly outline the two case histories of district CC services, before presenting an overview of the system dynamics model and the model-based experiments. Taking each case in turn, we then describe the insights obtained by the model-based analysis into the effects of the shifts in CC services (a base case analysis) and into how improvements could have been achieved (a policy analysis). Finally, we discuss the policy implications of our findings and highlight several areas for future research.

## **Two contrasting case histories of district CC services**

Prior to the introduction of district services, cardiologists from Ribsley and Veinbridge Hospitals developed and maintained their CC skills by using tertiary facilities. They both worked at the same tertiary centre. The level of both Ribsley and Veinbridge's CC activity was determined by the contracts between the individual tertiary centres and the Ribsley and Veinbridge purchasing

authorities. The elective demand for Veinbridge’s patients was met. However, there was a persistent problem of under-capacity for Ribsley’s patients. Consequently, temporary capacity increases had been provided at the tertiary level.

District CC services were originally introduced, using a mobile catheter laboratory, to compensate for the temporary closure of one of the tertiary-based catheter laboratories for repairs. Geographical considerations dictated that separate mobile laboratory sessions were required at Ribsley and Veinbridge Hospitals. At Ribsley, the district service sustained the capacity increases and thus produced dramatic improvements in access to elective services (Figure 1a). However, this trend was reversed when the tertiary laboratory reopened and the district service was withdrawn. A further temporary service was offered at a later date to compensate for a second period of construction work at the tertiary centre and to provide additional capacity. This achieved further temporary improvements in access.

INSERT FIGURE 1 NEAR HERE

At Veinbridge Hospital, the district CC service formed part of a long-term strategy to expand capacity and to develop cardiac services at Veinbridge. Consequently, after the tertiary laboratory reopened, the district CC service continued and was developed into a permanent service, with an integrated catheter laboratory.

The collaborators provided rich descriptions of the consequences of the shift in CC services. They described several different mechanisms by which the improvements in access stimulated demand. These mechanisms involved the influences of a number of factors, in particular, the average waiting time, knowledge of patients and GPs of CC and the new CC service, and expanding on the previous reports<sup>11</sup>, the skills of the CC operators. We considered the *rate* of the gain in skills and the *loss* in skills associated with staff turnover. We refer to these effects as the waiting time, knowledge and skills effects on demand. The stimulation of demand was also reflected in the hospital data. No other changes occurred during the 40-month observation period that could have accounted for the changes in demand.

In both cases, there was sufficient supply to meet the CC investigation waiting time goals whilst the district CC services were in place. However, in the Veinbridge case, the increase in demand for an outpatient appointment (see Figure 1b) produced severe detrimental effects on access to this service. This loss in access had not been anticipated.

### **The simulation model and experiments**

We provide an overview of the model and the experiments. We focus on the key model formulations because the model is too large to present in detail. The complete model documentation is available from the authors.

### *Model structure*

The simulation model was constructed using the *STELLA* software<sup>19</sup>. The model contains nearly 300 variables including 18 stocks (or levels) in the main structure and 52 overall. Stocks and their associated flows (or rates) are employed to: represent the physical flows in the system; calculate performance measures; and, compute the average values of several attributes following Homer's formulation<sup>20</sup>. To avoid cluttering the model with unnecessary detail, a well known modelling pitfall, we focused on changes to the processes of care rather than changes in location (Figure 2).

INSERT FIGURE 2 NEAR HERE

The model is divided into five main functionally related sectors or modules (Figure 3). It incorporates the collaborators' composite mental map of the mechanisms underlying changes in demand. The endogenous variables include several referral multipliers, the waiting list lengths, average waiting times and patient activity rates (numbers of patients diagnosed or treated). Beyond the model's explanatory boundary are a number of exogenous factors including the service capacities. This arose from the model purpose, which was to examine the consequences of the shifts in services and associated feedback effects and not the policy decisions to introduce a district service. Other exogenous variables reflected simplifying model assumptions. These included the base demand for an outpatient appointment that aggregated demand for new and follow-up appointments, and certain referral multipliers.

INSERT FIGURE 3 NEAR HERE

The modelling perspective of system dynamics is on continuous, aggregate phenomena. However, there have been calls to modify the system dynamics paradigm by incorporating discrete and stochastic elements into its models and by disaggregating further. This has been to extend the range of feedback phenomena that may be studied and to satisfy clients' desire for further detail<sup>21-23</sup>. In our study, it was necessary to make some modifications, for the former reason, as the emphasis was on decisions and processes at the local (individual hospital) level rather than the national level. This resulted in the introduction of discrete elements into the model including appropriate built-in *STELLA* functions. For example, to model the arrival and departure of trainee CC operators, a function was used which generated a pulse input of a specified size at a specified time. We also used a built-in logical function to specify the affordability limits for various costs including the district service running costs, district service preparation costs, procedural costs and overall costs. These limits represent the maximum amounts that could be afforded.

### *Calculation of the elective CC investigation referral rate*

The elective CC investigation referral rate is the rate of referrals from the inpatient route (from patients who were admitted into hospital and subsequently stabilised) added to the rate of referrals arising from patients seen in outpatient clinics. The former was assumed to be constant and the latter

is a fraction of the outpatient activity rate. This referral fraction is a reference referral fraction (described later) adjusted multiplicatively by several referral multipliers.

The model contains five base case referral multipliers. The first three are endogenous variables (Figure 4). The first multiplier involves the influence of the average waiting time (Figure 4a). This reflects how low waiting times could stimulate demand and how high waiting times could suppress demand. For the Veinbridge case, the waiting time was not considered a factor in referral decisions, which was consistent with the more ‘aggressive’ (confident and enthusiastic) referral behaviour at Veinbridge. This multiplier is set to a default value of 1 (zero effect).

The second multiplier concerns the influence of the knowledge of patients and GPs of the benefits of CC and the new CC service (Figure 4b). By developing a local CC service, GPs and patients became more knowledgeable about the benefits of CC and overcame anxiety about the risks, and thus more demanding for this service. The extent of this effect on demand increased as the district service grew generating more publicity and, through ‘word of mouth’, more reports of patients who had benefited. This effect was delayed as GPs and patients perceived changes in the availability of district services. Subsequently, this multiplier is modelled as a function of the perceived district CC rate. The effect was greater at Veinbridge because the CC service was permanent and heavily marketed.

#### INSERT FIGURE 4 NEAR HERE

The third multiplier depicts how changes in the skill base of those who select patients for a CC investigation could alter demand (Figure 4c). The effect of skills on demand is a weighted average of the skills multipliers of three categories of staff: expert CC operators; junior trainee CC operators; and, non-CC operators. Experts and non-CC operators were assumed to make referrals at constant rates with the former referring more than the latter due to their higher enthusiasm for CC and their greater skills in identifying patients in need (non CC operators will often be general physicians rather than cardiologists). The referral multiplier represents the changes in referrals of junior trainee CC operators as they climbed up the learning curve gaining specialist knowledge and confidence. The two functions reflect the different referral patterns, one representing Ribsley, involving a period of under-confidence and the other, representing Veinbridge, also involving a period of over-confidence. The multiplier reflects the greater confidence ‘spilling over’ during the learning process. Periodic skills effects were introduced by the existence of training programmes and the rotation of junior staff between hospitals.

The model contains two other referral multipliers that are, for simplicity, modelled exogenously. The first is the effect of significant capacity losses on referrals. It was assumed that a significant loss of capacity in the Ribsley case (e.g. the withdrawal of the district service) resulted in a reduction in referrals of the lowest priority, lowest risk cases as a ‘knee jerk’ reaction. For the Veinbridge case, as there were no significant capacity losses, this multiplier is set to a default value of 1 (a zero effect). The second multiplier is the effect of other factors on referrals. For the Veinbridge case, it reflects

how introducing a permanent CC service and opening the integrated laboratory prompted shifts of higher risk cases to the district level. For the case of Ribsley, this multiplier is set to the default value of 1, as the district service was only temporary.

The reference referral fraction represents periods where: there was no district service; normal capacity levels existed at the tertiary level; all district screeners were fully skilled; and, there was neither stimulation nor suppression of demand due to the waiting time or knowledge effects. This would correspond with a situation for which GPs' and patients' knowledge of CC relied totally on that derived from a tertiary-based service. An expert estimate applies to the case of Ribsley and for the case of Veinbridge, this figure is subjected to a multiplier to account for the more 'aggressive' referral environment. The formulation chosen for the referral fraction reflects the assumption that the 'knee jerk' reaction to the capacity losses dominated over the effects of waiting times, knowledge and other factors on referrals for an elective CC investigation. This avoids double counting patient referrals<sup>24</sup>.

### *Basic mechanisms*

The basic feedback mechanisms may be summarised with the simplified causal loop diagram shown in Figure 5. This depicts three balancing (or controlling) feedback processes and three reinforcing feedback processes. Note that further balancing loops, not shown, also represent the process of activity adjustment.

INSERT FIGURE 5 NEAR HERE

In the Veinbridge case, loop B3 was inactive due to the absence of the waiting time effect on demand. Loop R1 was also inactive because the expert CC operator made all final decisions about referrals and thus correcting any tendency to under- or over-refer. By contrast, the expert operator at Ribsley delegated more responsibility to the junior operators, and this included making all the final decisions about the patients they saw. The fact that the Veinbridge expert performed CC investigations *and* CC treatments might explain his decision to hold a tight leash over referral decisions.

### *Model parameterisation*

The justification underlying the model parameters varied as efforts were made to utilise all the available data, as recommended by Homer<sup>25</sup>. The approximations and assumptions made served for the purpose of the model. A number of parameters arose from archival data provided by the collaborative centres including various affordability limits, CC capacities and the average procedural costs. Other parameters were based upon estimates given by the study collaborators. They included the various response times, district facility preparation times and the nonlinear functions for the waiting time, knowledge and skills effects on demand. There was insufficient numerical data to

estimate these functions econometrically, as described by Sterman<sup>26</sup>. Having been given information about the basic shape and extremities, we drew smooth lines to construct the functions.

In several cases where numerical data did not exist, parameter values were derived from simple calculations from the actual activity rates and by assuming that parts of the system were in equilibrium and the desired waiting time goals were maintained. This referred to the reference outpatient rate and initial outpatient waiting list length for the Ribsley case.

The calibration of several Veinbridge parameters resulted from the descriptions of equilibrium and the assumption (made because the expert estimate proved to be unreliable) that the cardiologists there referred 50% more patients than those at Ribsley. This produced a calculated outpatient rate of similar magnitude to that of the Ribsley case. The outpatient rate could not be verified but served for our purposes. These parameters included the initial outpatient waiting list length and outpatient capacity level.

Some parameter values were derived from preliminary simulation runs. For example, the model was run for an initialisation period to set the average time spent on the waiting lists to the values observed in the real system. Further examples involved the use of sensitivity analysis and partial model testing. The former was used to derive the appropriate magnitude of the effect of other factors on demand to produce the reported capacity utilization rates for the Veinbridge case. The latter involved the magnitude of the effect of significant capacity losses on demand for the Ribsley case.

#### *Model testing procedures*

Confidence was gained in the model by applying a series of recommended and accepted tests for system dynamics models<sup>24,27</sup>. One key test considered the ability of the model to replicate the patterns of historical behaviour. In constructing the graphs of historical behaviour, known in system dynamics as reference modes, real data was only available for the CC variables in the Ribsley case. The other reference modes were constructed using hypothetical data based upon information provided by the collaborators. There was a good qualitative fit in both cases which was consistent with the aim of system dynamics modelling on reproducing the modes of behaviour (the basic trend) and avoiding the pursuit of point predictions<sup>15,28</sup>. Figure 6 compares the simulated and actual data for the Ribsley CC variables and Table 1 provides a number of historical fit summary measures, as described by Sterman<sup>29</sup>.

INSERT FIGURE 6 NEAR HERE

INSERT TABLE 1 NEAR HERE

Note that whilst the  $R^2$  statistic is low for the referral rate (Figure 6c), it is significantly different from zero. The bias (overestimates) in the average time spent on the CC investigation waiting list (Figure 6a) is due to two factors. Firstly, the early underestimates of the referral rate (Figure 6c). Fewer patients are considered to have just joined the waiting list and thus contribute, to the average



waiting time, a waiting time of zero. Consequently, higher average waiting times are estimated. Secondly, the underestimates of the tertiary-based CC investigation rate (Figure 6f). Estimating fewer waiting list removals produces higher waiting times. The bias in the CC investigation waiting list (Figure 6b) is due to the net effect of these two factors where the former dominates over the latter. The estimates of the tertiary-based CC investigation rate (Figure 6f) result from the estimates of the tertiary capacity. These were based upon approximations of the historical activity data and the reports that the system operated at full capacity.

### *Model-based experiments*

For each case, the model was calibrated, initial conditions were set and a base case simulation was produced. Various experiments were carried out to generate insight into the base case scenarios. A series of policy experiments followed which showed how the senior health professionals *could have* effectively intervened to improve the provision of services. In particular, we investigated the effectiveness of capacity changes and the use of stricter referral guidelines. Sensitivity analysis confirmed the insights generated by the various experiments. In this paper, we present the results of the key policy experiments.

Efforts were made to evaluate and interpret the experiments in broad terms. To address both the localised and broader consequences of the shift in CC services, both the supply and demand for services were considered across the cardiac referral chain. The key outputs were: the cumulative referrals, cumulative patient activity, cumulative overall costs, and the basic trends displayed by the average waiting times and waiting lists. We modelled how long patients currently on the waiting list have been waiting so far. Although this underestimates the actual time spent waiting, it served for our purposes as a performance measure given our focus on: (a) the basic waiting time *trend* (both waiting times would display similar trends) and (b) *comparisons* between policy runs and the base case.

Waiting times and waiting lists were both considered as each reflect different aspects of pressure on the system, respectively, the delay for services and the number of patients waiting. In the U.K. health system, the emphasis has switched between waiting lists and waiting times. In our case studies, the focus was on waiting times. The hospital consultants wanted to avoid excessive treatment delays and thus reduce patient anxiety and minimise the risk of patients deteriorating. As waiting time targets were specified in contracts for health services, hospital managers and purchasers were also concerned about reducing excessive treatment delays so that contractual obligations were met. However, those concerned about costs were also interested in the length of the waiting list. For example, if the waiting time target was met, an increase in the waiting list would indicate that increases in patient activity were necessary to prevent the average waiting time rising above its target.

In interpreting the results of the experiments, the needs of the different senior health professionals were considered. Firstly, the need to improve health i.e. identify patients in need of treatment, provide

health care promptly and appropriately, direct resources towards the most urgent cases, and increase activity and thus meet higher activity targets. Secondly, the need to control the overall costs incurred i.e. whilst agreeing to the district service set up costs, ensuring that services were used appropriately. Thirdly, the need to improve efficiency i.e. deliver care at the lowest cost/case. We thus refer to *health improvement, overall cost control and efficiency improvement perspectives*.

To summarise the graphical output of different simulation runs, for several variables, a simple ‘pressure summary index’ was defined by the area between each graph and a tolerance level as specified by the relevant goal (Figure 7). We assumed that pressure would only be exerted if the goal were exceeded. Comparing areas thereby quantified the degree to which improvements had been made over the duration of the simulation run, accounting for short-term and longer-term effects and changes in expectations, which would be reflected by adjustments in the tolerance level.

INSERT FIGURE 7 NEAR HERE

## **Ribbsley experiments**

### *The base case simulation*

The Ribbsley base case simulation was produced over a period of 54 months. Numerical data was available for the first 37 months. We then extrapolated over 17 months by assuming that no further district services, and no increases in tertiary capacity were provided during this period and that the referral behaviour of the Ribbsley cardiologists remained unchanged. Although no further numerical data was collected, contact continued with collaborators during this period and nothing occurred to contradict these assumptions. Given the persistent capacity shortages at Ribbsley, further temporary improvements in capacity would have been expected to occur at some time in the future. However, information about their timing and extent was not available. Therefore, in generating the base case scenario for Ribbsley, it was deemed inappropriate to extrapolate too far into the future.

Insight into the basic causes of the base case scenario was gained via a series of partial model test simulations and sensitivity analyses. These experiments demonstrated that the pressure on Ribbsley’s elective CC services was such that controlling the desired waiting time goal (see loop B1 in Figure 5) relied heavily on capacity changes (exogenous factors). The high pressure was maintained by the system being pushed harder when the district service was present (by seeking a lower waiting time goal), and by demand increasing. The latter occurred via several mechanisms including the waiting time (loop B2) and knowledge (loop R2a and R2b) effects on demand, which stimulated demand. Another factor was the changes in the skills base (loop R1) which led to periodic drops in demand as novice trainees periodically replaced experienced operators and, additionally, reductions in the degree to which demand was suppressed as they gained skills and confidence. The gains in skills and knowledge were not self-reinforcing as, again, they were driven by capacity changes. The temporary

increase in demand for an outpatient appointment prompted higher levels of cardiology outpatient activity as attempts were made to maintain the desired outpatient waiting time.

Different processes interacted with one another thus generating further pressure on the system. For example, the increases in referrals for CC investigations in response to the reductions in the average waiting time drove the district activity to rise to control the waiting time. This generated more learning experiences for junior CC operators and thus reinforced the gain in CC operator skills (interactions between loops B1, B2 and R1). In spite of the increasing pressure, the capacity changes produced significant reductions in the CC waiting list and waiting times. However, these reductions were only temporary.

The increase in demand for outpatient services was maintained for a short period beyond the withdrawal of district services in month 23 due to the inherent perception delay of GPs and patients (loop R2b). The outpatient waiting list and average waiting time exhibited small temporary increases. Eventually the referral expectations dropped and equilibrium was re-established. Also due to the perception delay, when the district service was reintroduced in month 34, it was in place for an insufficient length of time to prompt further increases in demand.

#### *Policy analysis: Seeking improvements in service provision*

The Ribsley policy experiments indicated that the extent of the under-capacity was such that demand management strategies alone (diminishing the strength of loops B2, R2a and R2b), even the use of the most stringent clinical guideline, could not have altered the undesirable rise in the CC waiting list and average waiting time. Frequent capacity increases were necessary (to intensify loop B1). Furthermore, coordinating capacity increases with efforts to manage demand would ensure that the benefits of increasing supply were not cancelled out by stimulated demand.

Different approaches to capacity increases would have produced different effects (Figure 8 and Table 2). An obvious approach to increasing capacity would have been to provide a permanent district service, as this would have been expected to maintain the access targets permanently. However, from an *overall cost control perspective*, the benefits of increasing the supply would have been cancelled out by stimulated demand. The CC investigation waiting list would have exhibited a gradual rise thus indicating the need for further increases in the elective CC investigation rate in order to maintain the desired waiting time. Furthermore, the stimulated demand would have created a new problem, as outpatient capacity shortages would have arisen. This would have been unanticipated and it would have then called for further resources in an attempt to control access to outpatient services. From a *health improvement perspective*, the reduction in the CC waiting times and increases in activity associated with a permanent district service would have been attractive, but these benefits would have been undermined by the loss in access to outpatient services.

INSERT FIGURE 8 NEAR HERE

## INSERT TABLE 2 NEAR HERE

By using new referral guidelines, it might have been possible to suppress demand and generate considerably lower costs. However, whilst the use of new moderate guidelines would have enabled the access to CC services to be maintained, strict new guidelines would have been necessary to eliminate the outpatient capacity shortages. The feasibility of introducing strict new guidelines would have been doubtful even in a modest referral environment such as Ribsley. Therefore, even with realistic safeguards in place to control demand, a permanent district service could not have led to significant improvements from either an *overall cost control* or *health improvement perspective*.

### *Limiting the use of district services*

The access problems generated by a permanent district service suggested that a more effective approach to increasing supply would have been to limit the use of district services. In theory, this could have involved either expanding the tertiary-based service, and just using the district service to compensate for tertiary facility closures, or offering further temporary district services. Compared to a permanent district service, carrying out the same capacity increases at the tertiary level would have generated fewer costs. This is because limiting the use of district services would have restricted the degree to which demand was stimulated (by limiting the knowledge effect on demand). Less demand would have meant fewer patients would have needed to be treated in seeking to maintain the desired waiting time. Therefore, this option would have been more attractive from an *overall cost control perspective*. However, in practice, expanding the supply for elective services at the tertiary level would have been difficult whilst meeting the demands for more urgent cases.

Expanding the supply at the district level would have been more feasible, as the service would have been solely devoted to elective care. Those pursuing an overall cost control agenda would have favoured further temporary district services over the same overall capacity increases translated into a permanent expansion at the tertiary level. This would apply because the former would have generated lower costs due to the ‘knee jerk’ reductions in referrals in reaction to the capacity losses. Note that this experiment was supposed to reflect a continuation of the base case scenario in that the capacity increases were unexpected. If further capacity increases were expected, then it could be assumed that the ‘knee jerk’ reaction would not apply, as the cardiologists would increase their referral rates in anticipation of the next capacity increase.

The preference from a *health improvement perspective* would have been less clear. The greater use of a district service would have provided more opportunities to devote tertiary resources to more complicated cases. It would also have led to the stimulation of more demand for outpatient services. Bringing more patients forward for assessment could have led to the identification of further high-risk patients and also supported higher activity targets. However, fewer referrals would have been made for CC services, as by introducing more capacity losses there would have been more ‘knee jerk’

reductions in referrals in reaction to these losses. It could be assumed that these reductions would have referred to lower risk patients. However, their assessment as lower risk patients would have been based on incomplete information i.e. without the benefit of the CC investigation, which was the most accurate diagnostic tool available. It would have been possible that some high-risk cases that presented minor symptoms would have slipped through the net. Therefore, from a health improvement perspective, further temporary district services would have involved trade-offs. Nevertheless, it could be assumed that those pursuing this agenda would have conceded that, on balance, further temporary district services would have been the only practical way to achieve improvements in access.

The attractiveness of further temporary district services would have been subject to district services being efficient. Based upon the limited data provided, the district service provided investigations at a lower cost/case compared to tertiary-based investigations. However, some patients needed to undergo a second CC procedure for treatment as synchronous investigation and treatment was only permitted at the tertiary level. Taking this factor into account, a district service would only have been attractive from an *efficiency improvement perspective* if it had avoided a high proportion of patients undergoing their CC investigation as an inpatient.

## **Veinbridge experiments**

### *The base case simulation*

The Veinbridge base case simulation was produced over a period of 72 months (Figure 9). This included model projections of up to 27 months, based on the assumptions that the capacities for outpatient and elective CC investigation services remained constant and that there were no further shifts to the district level. Contact with the collaborators continued until month 53 and nothing occurred to contradict these assumptions.

INSERT FIGURE 9 NEAR HERE

The base case experiments demonstrated that prior to the introduction of district CC services to Veinbridge (at month 13), the system was stable and the waiting time goals were met (Figures 9a and 9e). Furthermore, neither the outpatient service nor the elective CC service operated at full capacity. This indicated that there was some scope to raise activity levels if desired (see loop B1 in Figure 5). The drop in the tertiary-based CC investigation rate, associated with the closure of one of the tertiary labs, produced a shift in elective CC investigations to the district level (loop B1) and quickly pushed the remaining tertiary-based service to operate at full capacity (Figure 9d). These processes acted to ensure that the desired waiting time was maintained. The adjustment in district activity was also prompted by demand being stimulated (Figure 9c). The consultant cardiologist took advantage of the district service to investigate more borderline cases (an exogenous factor). The development of the district services led to increased knowledge about CC among patients and GPs and this stimulated

demand for both outpatient and CC services (loops R2a and R2b). The increase in district activity reinforced the effect of knowledge on demand.

By the time demand was stimulated by GPs and patients, the tertiary service had already been driven to operate at full (albeit reduced) capacity. Therefore, the process underlying the knowledge effect on demand for elective CC investigations (loop R2a) did not reinforce the adjustments in tertiary-based CC investigation rates. By contrast, more district activity generated more knowledge about CC which stimulated demand and thus led to calls for more activity in order to maintain the desired waiting time (interactions between loops B1 and R2a). Further interactions were produced as further demand for outpatient services was also generated and this led to calls for more outpatient activity to maintain the desired outpatient waiting time. The increases in outpatient activity produced further referrals for CC investigations and this contributed to calls for more district activity (interactions between loops B1 and R2b).

As the district service was not operating at full capacity, there was scope for further increases in the district investigation rates (Figure 9d). This occurred when the integrated catheter laboratory opened at Veinbridge and the consultant cardiologist took advantage of the opportunity presented by the new facility to investigate more borderline cases. This prompted the district activity rate to increase (loop B1) in order to maintain the desired waiting time goal.

The CC investigation waiting list increased (Figure 9b) and thus required the elective CC investigation rate to rise to maintain the desired waiting time. The waiting list rose due to the inherent delay between the activity rates adjusting (loop B1) in response to the increase in referrals (loop R2a), and the fact that efforts were only made to adjust activity to meet a desired waiting time and not a desired waiting list length. In general, seeking a desired waiting time will meet and maintain this goal provided there are sufficient resources available to respond to changes in the waiting list length with the necessary changes in the waiting list removal rate. However, maintaining the average waiting time at its desired level will not necessarily indicate that the waiting list goal is simultaneously met. The waiting list could, in fact, be elevated or rising above its desired level. Therefore, capacity increases could not have prevented the CC investigation waiting list rising.

The outpatient waiting list and average waiting time (Figure 9e) exhibited steady increases due to capacity shortages. These variables displayed asymptotic growth due to the action of a balancing process that controlled other waiting list removals (loop B3). It only represented a minor feedback process whilst other processes were active, but as the saturation points were reached this loop became dominant. The action of this loop caused the outpatient waiting list removals to rise towards the level of the waiting list referrals.

### *Policy analysis: Seeking improvements in service provision*

The Veinbridge base case analysis indicated that whilst the long-term strategy to expand CC services at Veinbridge Hospital relied upon increases in demand, the strategy would be undermined unless it was coordinated with controls to limit the stimulated demand. Demand for outpatient services (where patients are screened for a CC investigation) would have to be controlled. Significant increases in the Veinbridge outpatient activity were not essential to support the expansion of CC services at Veinbridge. Demand for the CC service also arose from elsewhere (cardiologists from the surrounding districts also used the Veinbridge facilities). The Veinbridge purchasers were concerned that the increases in demand for outpatient services indicated a rise in inappropriate referrals. Consequently, they were unwilling to provide unlimited funding for outpatient services.

### *Combining controls on demand with new forces driving activity*

The base case analysis also suggested that with sufficient slack in the system, the waiting time and waiting list goals could have been met simultaneously by changing the forces that drive activity rates - seeking a desired waiting list length rather than a desired waiting time. There was spare capacity for elective CC services and introducing controls on demand (diminishing the strength of loop R2b) would have produced some spare capacity for outpatient services.

The policy analysis thus considered the use of stricter clinical guidelines in combination with changes to the goals that drive activity. The use of stricter guidelines was modelled by assuming that the maximum degree to which demand could be stimulated was cut by 50%. This was considered to form a reasonable adjustment. The factors limiting a further reduction were the acceptability of such guidelines in such an 'aggressive' referral environment, and for the case of CC services, the need to ensure sufficient demand to justify the development of a permanent district service at Veinbridge. For comparative purposes, a further experiment considered the use of stricter guidelines alone, and another experiment investigated the effects of a 10% increase in outpatient capacity.

The experiments demonstrated that, the outpatient and CC waiting list and waiting time goals could have been simultaneously (and feasibly) met only by using stricter clinical guidelines in combination with seeking a desired waiting list length (Figure 10 and Table 3). In fact, this policy would have overcompensated by producing an average waiting time that was lower than required. The use of stricter new referral guidelines alone would have led to some improvements but fewer than those derived from the combined policy. Increasing outpatient capacity would have generated increases in the overall costs and CC costs in particular as further patients were pushed along the referral chain.

INSERT FIGURE 10 NEAR HERE

INSERT TABLE 3 NEAR HERE

### *Achieving significant and sustainable improvements*

Therefore, the implementation of the long-term strategy to expand CC services could have been improved by coordinating the shift in CC services with the use of new referral guidelines and changes to the forces that drive activity rates. This would have produced significant and sustainable improvements in behaviour that would have been attractive from both an *overall cost control* and a *health improvement perspective*. The effects of this policy would have been particularly attractive from an overall cost control perspective, as it would have led to improvements in access and reductions in costs. For those pursuing a health improvement perspective, there would have been trade-offs. They would have valued the improvements in access to outpatient services and reduction in the CC investigation waiting list but the reductions in referrals and activity would have been in conflict with the desires to meet higher activity targets and identify more high risk patients.

As with the case of Ribsley, taking into account the need for some patients to undergo further CC, a mobile-based district service at Veinbridge was only more *efficient* than a tertiary-based service if it avoided a large proportion of patients being catheterised as inpatients. When the integrated catheter laboratory at Veinbridge opened, it provided the opportunity to improve the efficiency of the district service as the cost/case declined with the volume of patient activity. This eliminated any conflict there might have been in considering both a *health improvement* and an *efficiency improvement perspective*.

## **Discussion**

Little attention has been given to the actual mechanisms of feedback effects associated with service shifts. This is in spite of the increasing emphasis in health care on the need for ‘joined-up’ or ‘whole systems’ thinking where the different parts of the system and their interactions are considered<sup>12,30,31</sup>. This paper has demonstrated the contribution system dynamics modelling can provide. Our study offers a plausible causal framework to support the hypothesis that shifting services can stimulate demand. Also, in offering empirical evidence for this phenomenon, we argue that such policies should be coordinated with other policy changes to control demand and to ensure that suitable forces drive patient activity.

As with all models, this model is based upon certain simplifying assumptions. For example, in reality, there would be additional feedback mechanisms. Nevertheless, the model still successfully replicated the problematic behaviour reflected in real world data for both cases. Furthermore, the findings of this research are not specific to the subject, timing or location of the case studies as the interactions between supply and demand are widely generalisable. Therefore, a number of broad policy lessons and recommendations may be derived. These are as follows:



- *Increasing capacity is not necessarily the most effective way of improving access*

It has been previously argued by Wolstenholme<sup>30</sup> in a community care setting, that increasing capacity is not necessarily the most effective way of improving access. Our study provides further evidence of this paradox. Wolstenholme highlights the greater leverage of flow variables compared to that of stock variables. The Veinbridge case showed that capacity increases (a stock variable) and stricter referral guidelines (flow variables) would have provided similar leverage in improving behaviour in terms of reducing the waiting list and average waiting time, but significantly better leverage could have been obtained by combining stricter guidelines with changes to the goals that drive activity (both flow variables). This result arose from the existence of spare capacity and was not specific to the case of the shift in CC services. It was a general consequence of the interplay between supply and demand variables that determine the waiting list length and average waiting time. Although discussions of spare capacity may seem incongruous with reports of long NHS waiting lists and waiting times, some spare capacity is often released by service shifts, by their ability to provide additional capacity and/or by prioritising patients. By contrast, the Ribsley case demonstrated that in cases of extreme imbalance between supply and demand, only capacity increases could have provided the necessary leverage. However, demand management strategies could still have played an important role by improving the leverage of capacity increases.

- *Focusing on isolated events, short-term results and single performance measures can lead to ineffective policies and misleading conclusions.*

Our study challenges the persistent tendency in health care towards a narrow focus on isolated events, short-term results and single performance measures with calls made to shift the emphasis from the waiting list length on to the waiting time<sup>11</sup>. For example, we showed that a permanent district service at Ribsley would have only produced similar short-term qualitative improvements in access to CC services, compared to an expanded tertiary-based service. In addition, the former would have slowly produced access problems further down the referral chain at outpatient services. Another example arises from the Veinbridge analysis. It showed how maintaining a waiting time goal did not necessarily mean that the system was free from pressure; the rise in the waiting list suggested that it was under pressure and that the waiting time goal was only being maintained because more money was being poured in to raise activity levels.

- *Patient pressure represents challenges to the delivery of health care which cannot be ignored in a realistic model*

The case studies also illustrated the influence of pressure by patients on clinical decisions and the problems that can arise from the inability to cope with this pressure and the poor management of demand. It has been argued that clinical and policy decisions should be driven by the preferences of

patients and the public<sup>33,34</sup>. Questions thus arise about how the shift in the balance of care can continue whilst providing high quality care to patients whose expectations for health care are traditionally high.

### *Generalisations*

The model represents a plausible hypothesis of the effects of shifts in health services to improve access and it was parameterised to two cases of the shift in CC services. Variable names that referred specifically to the shift in CC services were used to ensure that the model had face validity, but the variables could be relabelled to more general terms.

Whilst the NHS formed the context to our case studies, our research findings have implications for health service shifts in other countries in spite of the differences between their health systems. The NHS is a publicly funded system. Other types of health systems are privately funded or insurance-based. The stimulation of demand in response to improved access is a common response in the NHS and it is a consequence to services being free at the point of delivery. However, it could be argued that services in other health systems, whether private or insurance-based, are all free to some extent in one sense or another. The individual feedback mechanisms that we have discussed can be generalised to other health systems. Even the impact of waiting times on referrals can be generalised in spite of the fact that waiting lists are a typical characteristic of the NHS; the waiting time can translate to other health systems into the price customers pay for services. The waiting time is a price in some form and its causal relationship with demand is the same.

### *The 'supplier induced demand' debate*

Our study contributes to the 'supplier induced demand' (SID) debate by offering evidence of the existence of inducement and by providing insight into the underlying feedback mechanisms. SID refers to a doctor acting as an imperfect agent by providing or recommending more (or less) services than a fully informed and knowledgeable patient would order<sup>3,4</sup>. In our study, we considered how patients demanded more services as they gained knowledge and thus led to increases in the utilisation of CC services. This would suggest that prior to the acquisition of this knowledge, both the GPs (in considering referrals for an outpatient appointment) and cardiologists (in considering referrals for a CC investigation) were acting as imperfect agents by suppressing demand. However, we should point out that our model aggregates the response to increased knowledge about both the health benefits of CC and the ability to undergo CC locally. Providing a local district service instead of a more remote tertiary service could be perceived as an improvement in service quality. Changes in demand that relate to changes in service quality alone do not support the existence of SID as they can be explained within the standard theory of supply and demand<sup>35</sup>.

The effect of other factors on demand as presented by the Veinbridge case and the waiting time effect on demand could involve doctors “altering the perceptions of the patient’s needs and of the capacity of medical technology to satisfy them”, which is a description of SID by Evans<sup>36</sup>. However, the latter effect would not indicate the existence of SID if patients considered the waiting time to be a key factor in determining their desire for a CC investigation.

### *Potential changes in health*

In both cases, the stimulated demand prompted increases in activity. The question of whether or not the increase in the CC investigation rate actually improved health was not fully addressed as serial data of clinical events and long-term patient outcomes were not monitored. However, several comments could be made. In the Ribsley case, the waiting time for catheterisation dropped and this contributed to shorter delays for those requiring invasive treatment (coronary angioplasty or coronary bypass surgery). This could offer the potential for health benefits as treatment delays can lead to some patients deteriorating.

The fraction referred on for a CC investigation increased in both cases. As the case mix of the patient populations did not change, this meant that the threshold for a CC investigation had changed towards less severe cases (defined without the benefit of a CC investigation). Catheterising more of these patients could lead to improvements in health if it identified further patients in need of invasive treatment; some patients with severe heart disease only display minor symptoms and these patients may thus fail to be identified by other diagnostic methods. The fraction of patients referred on for invasive treatment remained constant so, associated with the increased catheterisation rate was an increased invasive treatment referral rate. This would suggest that, from the less severe cases, further patients in need of such treatment had, indeed, been identified. However, for even less severe cases, it might be better to delay bypass surgery until the disease is more advanced as repeat bypass surgery, which can occur given the progressive nature of heart disease, carries higher risks. Therefore, delaying catheterisation (and therefore delaying bypass surgery) could, paradoxically, be beneficial in the long-term management of the disease.

### *Further work*

The research described in this paper could be developed in several different ways. The model could be employed to explore other service shifts, of which there are numerous examples. This would serve to verify if the research findings could be replicated, and could thus increase the evidence base that improving access can stimulate demand, or if the findings warranted some revision. The model boundary could be extended to endogenise the follow-up process of patients after their discharge from an outpatient appointment and the process of changes in the referral threshold for CC and district CC as the district service evolves. The model could also be disaggregated to elucidate the dynamics of

changes in clinical priority between routine and urgent elective cases. Finally, the pressure summary index could be investigated further to establish its scope and limitations and thus draw up recommendations for its use and, if necessary, its refinement.

### *Acknowledgements*

We would like to thank all those who collaborated in this study, for their time and input, and The Wellcome Trust for their sponsorship of this research (Reference Number 041243).

### **References**

1. Coulter A (1995). Shifting the balance from secondary to primary care: needs investment and cultural change. *BMJ* **311**: 1447-1448.
2. Hensher M, Fulop N, Coast J, Jefferys E (1999). Better out than in?: alternatives to acute hospital care. *BMJ* **319**: 1127-1130.
3. Hay J, Leahy MJ (1982). Physician-induced demand: an empirical analysis of the consumer information gap. *J Health Econ* **1**: 231-244.
4. Rice T (1984). Physician initiated demand for medical services. *J Health Politics, Pol. Law* **14**: 587-600.
5. Roland M, Morris R (1988). Are referrals by general practitioners influenced by the availability of consultants? *BMJ* **297**: 599-600.
6. Newton JN, Henderson J, Goldacre MJ (1995). Waiting list dynamics and the impact of earmarked funding. *BMJ* **311**: 783-785.
7. Goddard JA, Tavakoli M (1998). Referral rates and waiting lists: some empirical evidence. *Health Econ* **7**: 545-549.
8. Hamblin R, Harrison A, Boyle S (1998). *Access to elective care: why waiting lists grow*. London: King's Fund.
9. Yates J (1987). *Why are we waiting?* Oxford: Oxford University Press.
10. Harrison A, New B (2000). *Access to elective care: what should really be done about waiting lists?* London: King's Fund.
11. Hamblin R, Harrison A, Boyle S (1998). The wrong target. *Health Serv J* **108**: 28-31.
12. Scott A (1996). Primary or secondary care? What can economics contribute to evaluation at the interface? *J Public Health Med* **18**: 19-26.
13. Godber E, Robinson R, Steiner A (1997). Economic evaluation and the shifting of balance towards primary care: definitions, evidence and methodological issues. *Health Econ* **6**: 275-294.
14. Miller P, Graig N, Scott A *et al* (1999). Measuring progress towards primary care-led NHS. *Br J Gen Pract* **49**: 541-545.
15. Forrester JW (1961). *Industrial dynamics*. Cambridge (MA): MIT Press.
16. Taylor K, Lane D (1998). Simulation applied to health services: opportunities for applying the system dynamics approach. *J Health Serv Res Policy* **3**: 226-232.

17. Jun JB, Jacobson SH, Swisher JR (1999). Application of discrete-event simulation in health care clinics: a survey. *J Opl Res Soc* **50**: 109-123.
18. Taylor K. A system dynamics model for planning and evaluating shifts in health services: the case of cardiac catheterisation procedures in the NHS. PhD thesis, 2002, London School of Economics and Political Science.
19. High Performance Systems *STELLA software manuals 1985*. High Performance Systems, 145 Lyme Road, Hanover NH 03755, USA.
20. Homer JB (1987). A diffusion model with application to evolving medical technologies. *Technol Forecast Soc* **31**: 197-218.
21. Wolstenholme EF (1999). A patient flow perspective of U.K. Health Services: exploring the case for new “intermediate care” initiatives. *Sys Dyn Rev* **15**: 253-271
22. Crawford CM (1991). Endogenous safety processes: a model of regulation and safety in industrial firms. *Sys Dyn Rev* **7**: 20-40.
23. Scholl GJ (1995). Benchmarking the system dynamics community: research results. *Sys Dyn Rev* **11**: 139-155.
24. Coyle RG (1996). *System dynamics modelling*. London: Chapman and Hall.
25. Homer JB (1997). Structure, data and compelling conclusions: notes from the field. *Sys Dyn Rev* **13**: 293-309.
26. Sterman JD (2000). *Business Dynamics*. New York: Irwin McGraw Hill.
27. Forrester JW, Senge PM (1980). Tests for building confidence in system dynamics models. In: Legasto AA, Forrester JW, Lyneis JM (eds). *TIMS Studies in the Management Sciences* **14**: 209-228.
28. Richardson GP (1991). *Feedback thought in social science and systems theory*. Philadelphia University of Pennsylvania Press.
29. Sterman JD (1984). Appropriate summary statistics for evaluating the historical fit of system dynamics models. *Dynamica* **10**:51-66.
30. Spurgeon P (2001). Development of clinical care networks. In: Merry P (ed). *NHS Handbook 2001/2002*. 16<sup>th</sup> edition. East Sussex: JMH Publishing: 191-193.
31. Smith R (1999). Reconfiguring acute hospital services. *BMJ* **319**: 797-798.
32. Wolstenholme EF, Coyle RG (1980). Modelling discrete events in system dynamics models: a case study. *Dynamica* **6** :21-27.
33. Kassirer JP (1983). Adding insult to injury: usurping patients’ prerogatives. *N Engl J Med* **308**: 898-901.
34. Hornberger JC, Habraken H, Bloch DA (1995). Minimum data needed on patient preferences for accurate, efficient medical decision-making. *Med Care* **33**: 297-310.
35. Ryan M, Mooney G (1992). *Supplier induced demand: where are we and where should be go?* Health Economics Research Unit. University of Aberdeen. Discussion Paper 05/92
36. Evans RG (1974). Supplier-induced demand: some empirical evidence and implications. In: M Perlman (ed). *The Economics of Health and Medical Care*. New York: John Wiley and Sons: 162-173.

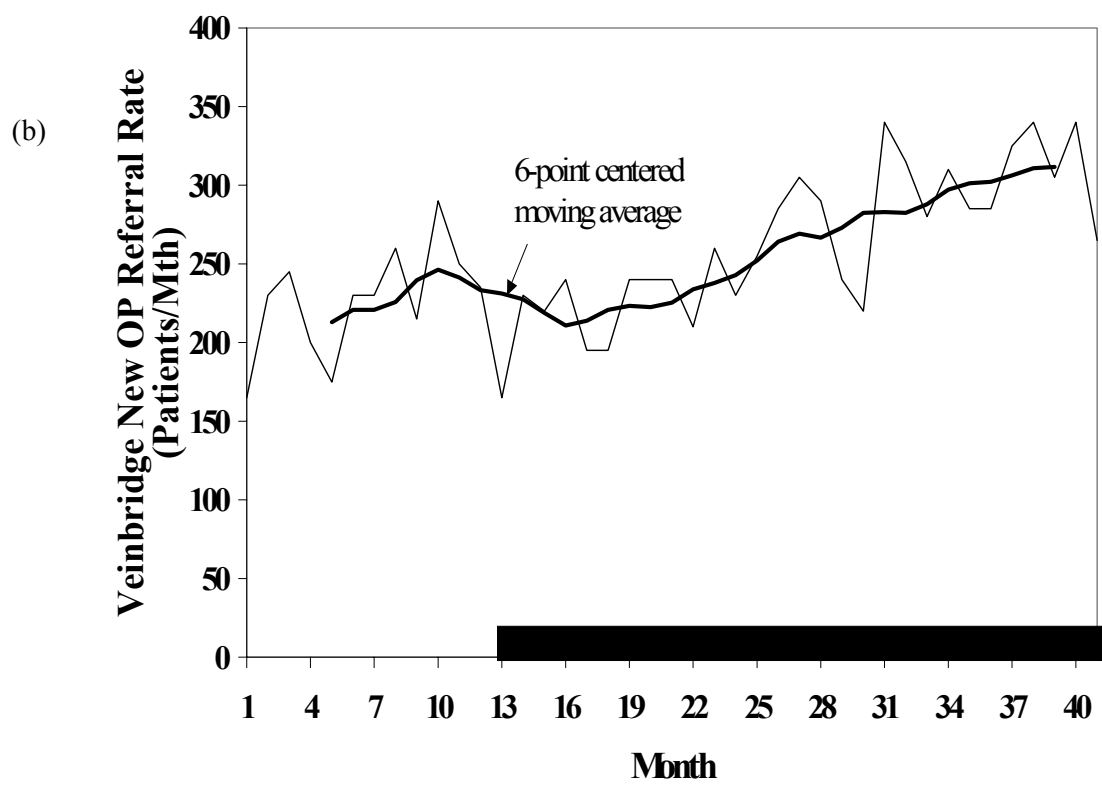
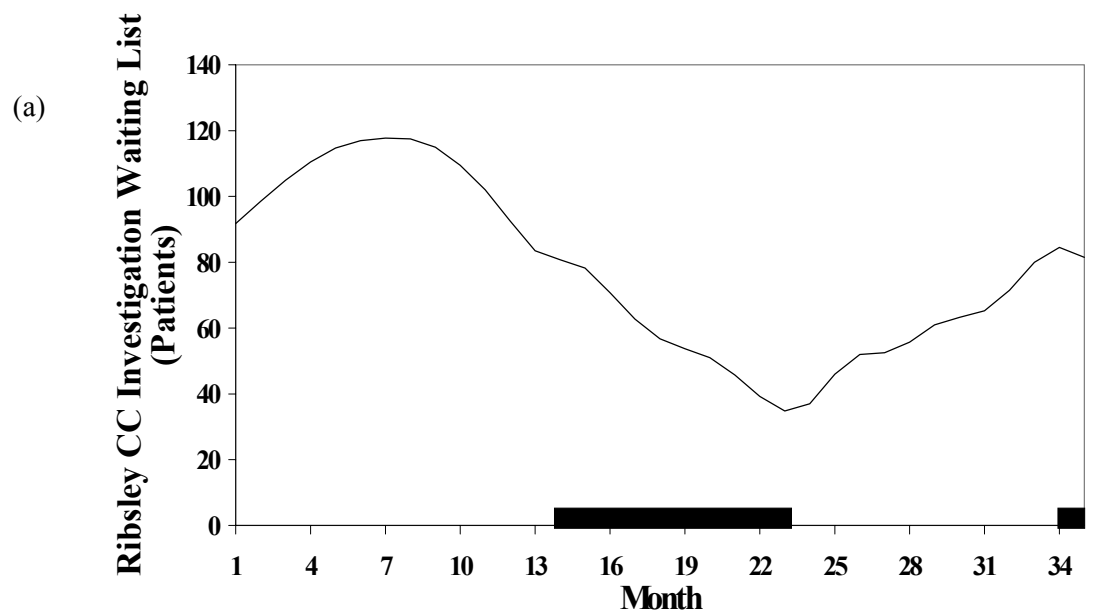


Figure 1

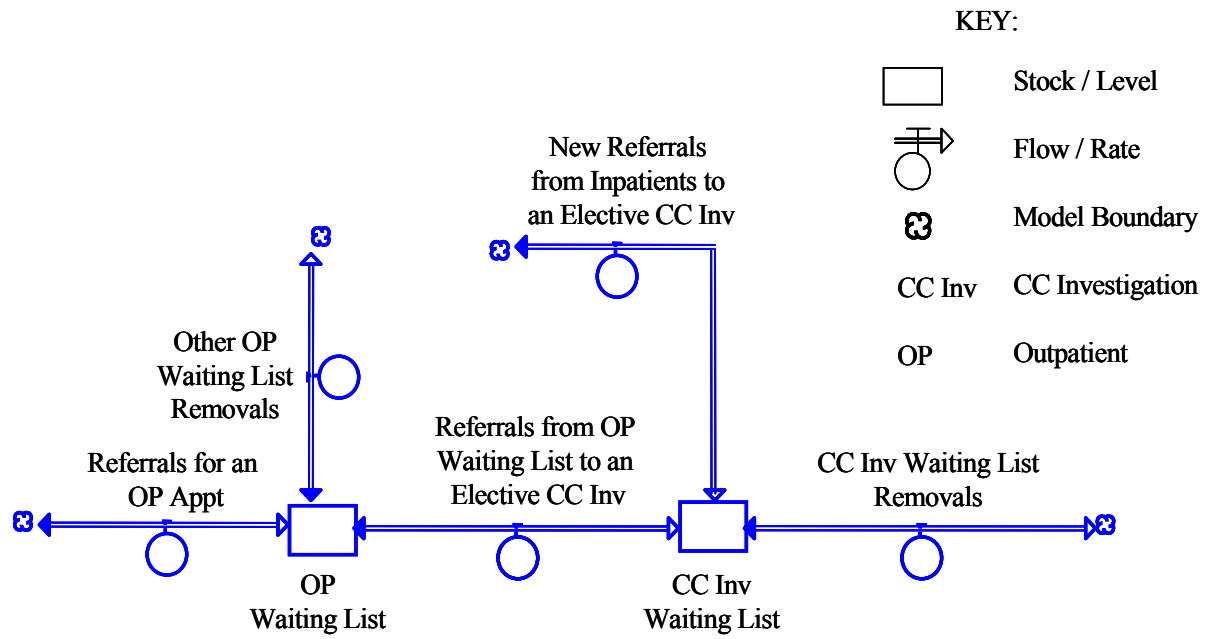
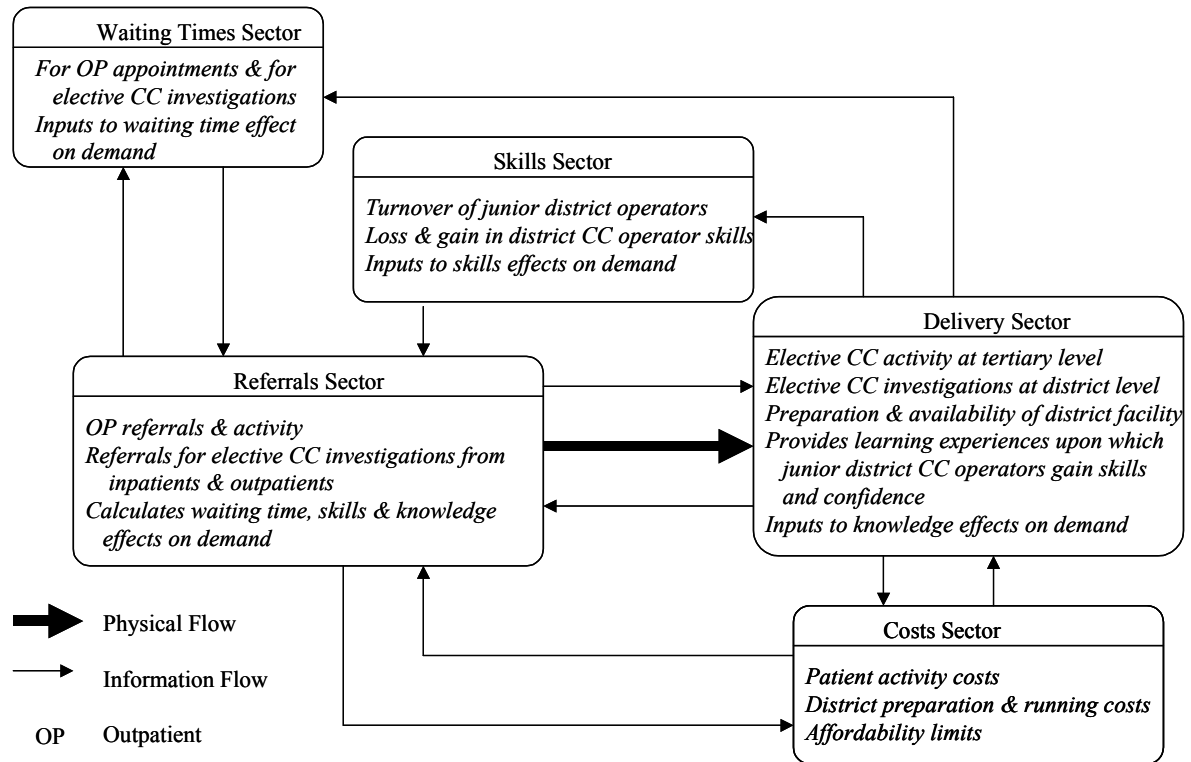


Figure 2



**Figure 3**



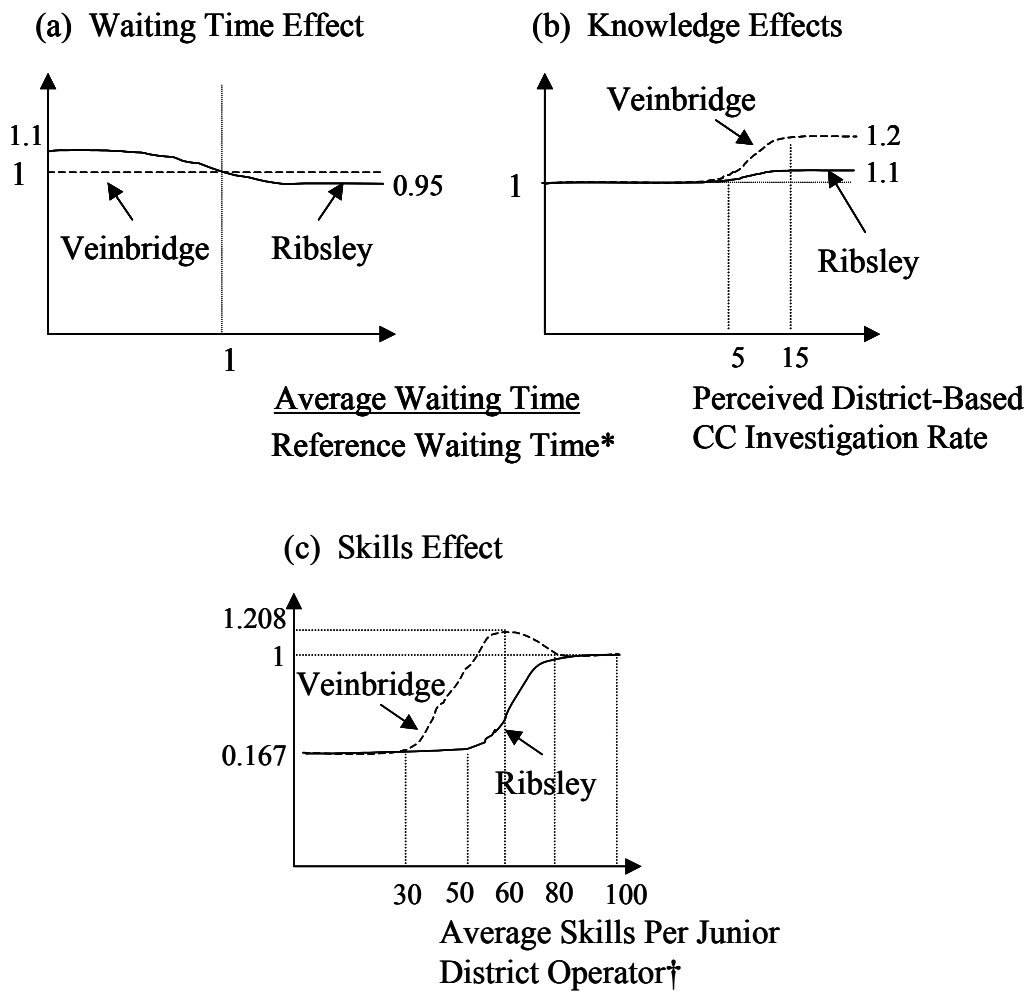


Figure 4

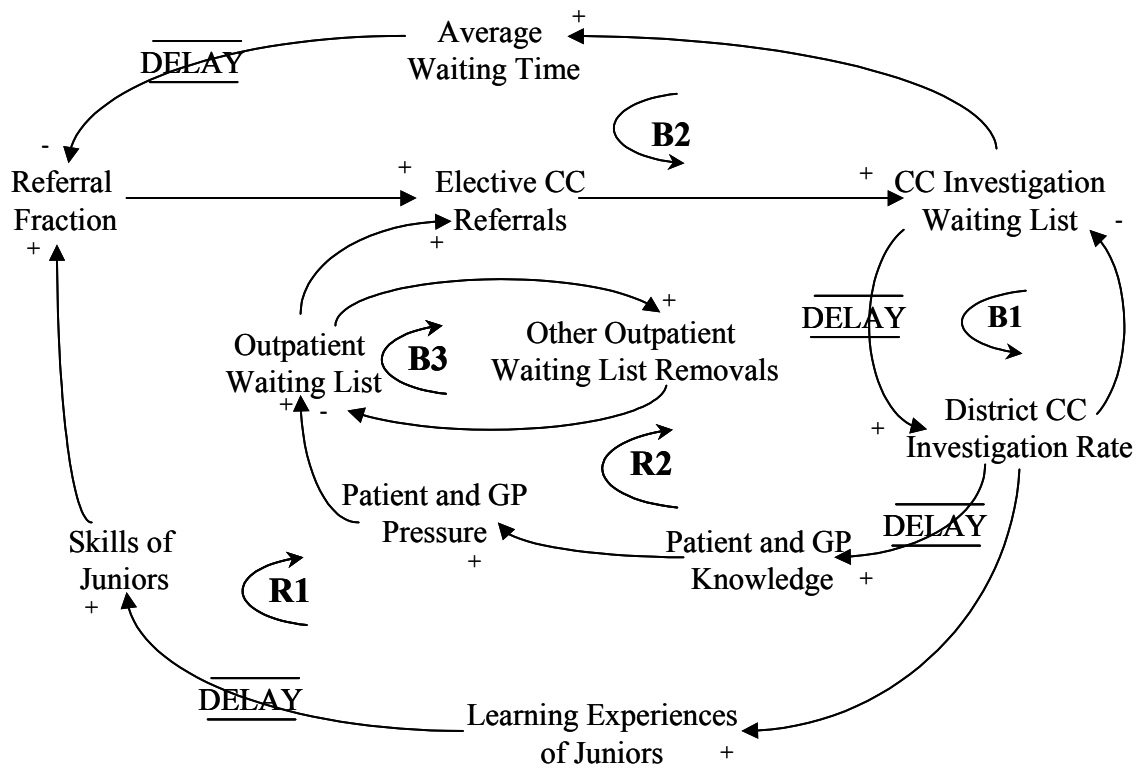
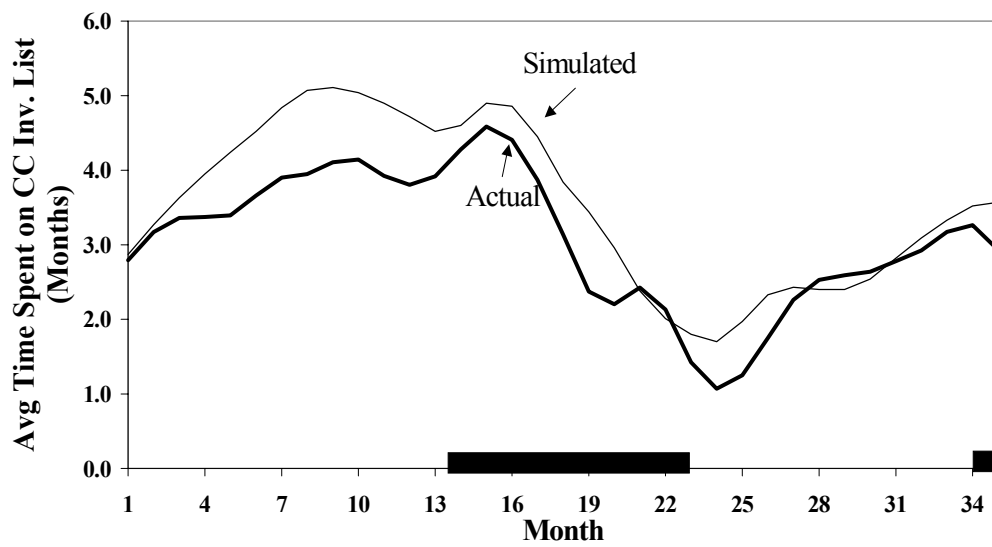


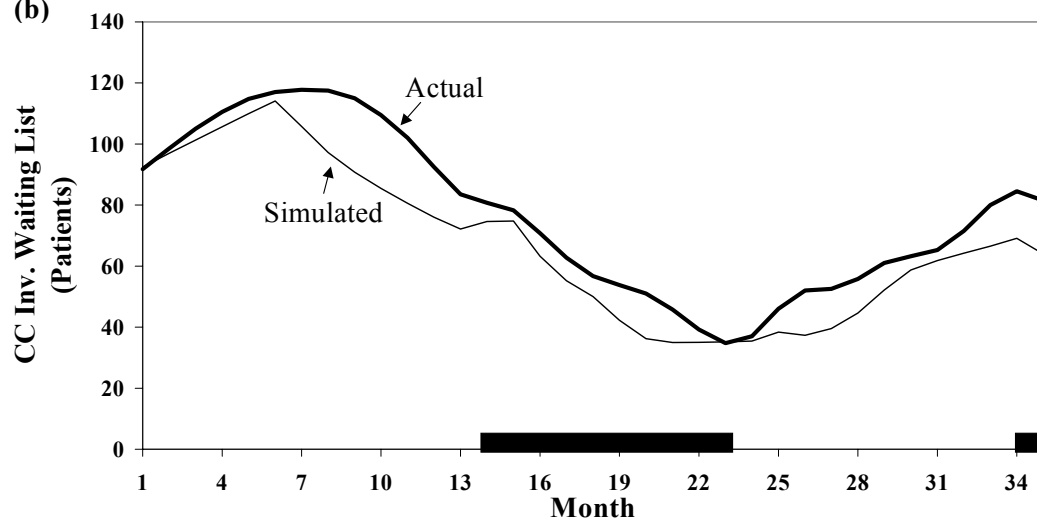
Figure 5

(Figure 6 has been enlarged and runs over 3 pages)

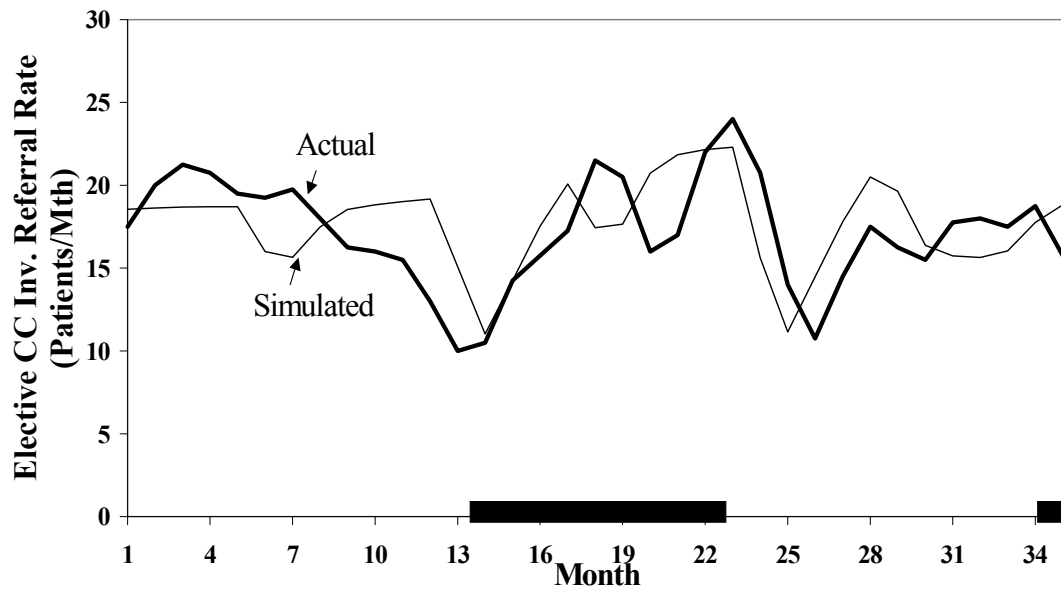
(a)



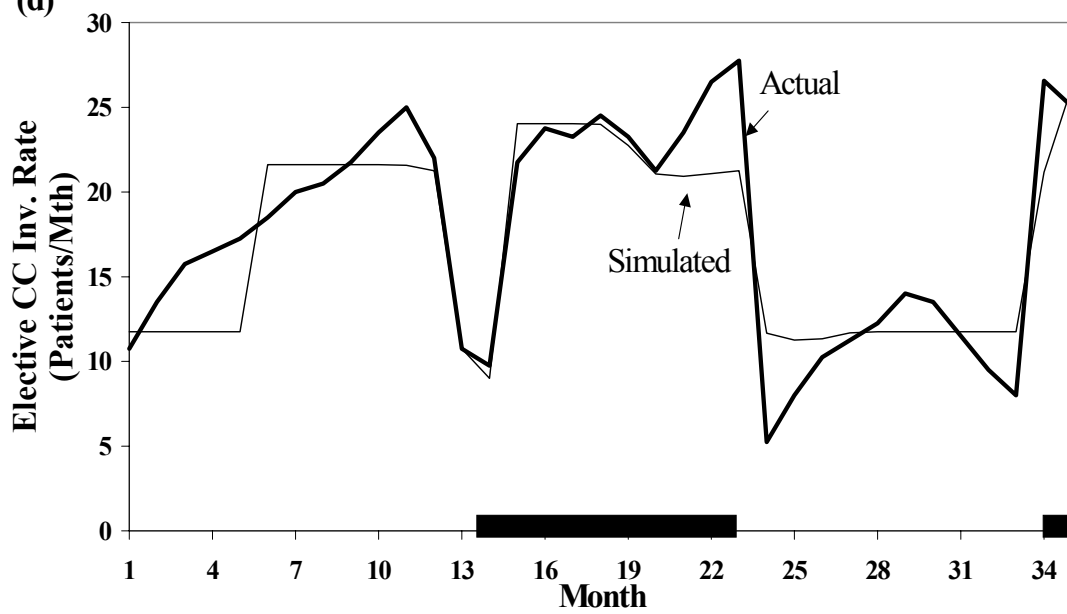
(b)



(c)



(d)



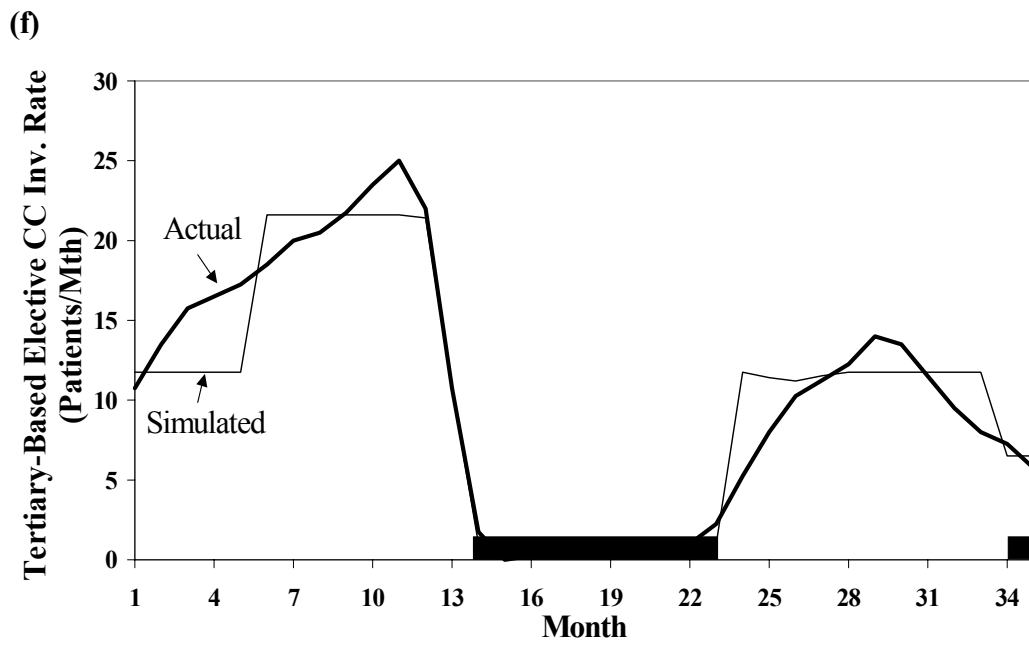
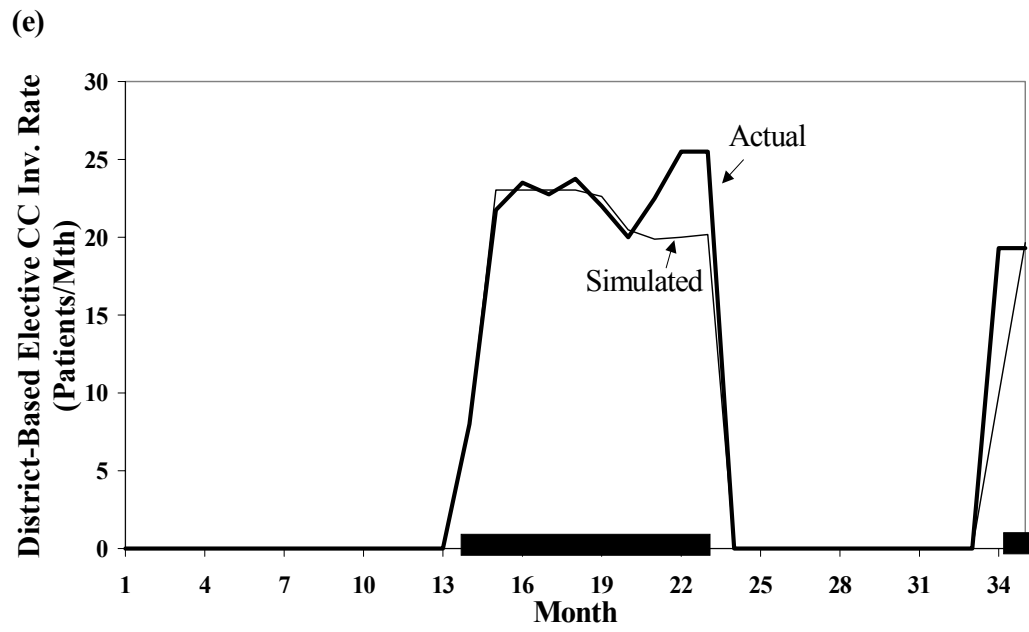


Figure 6

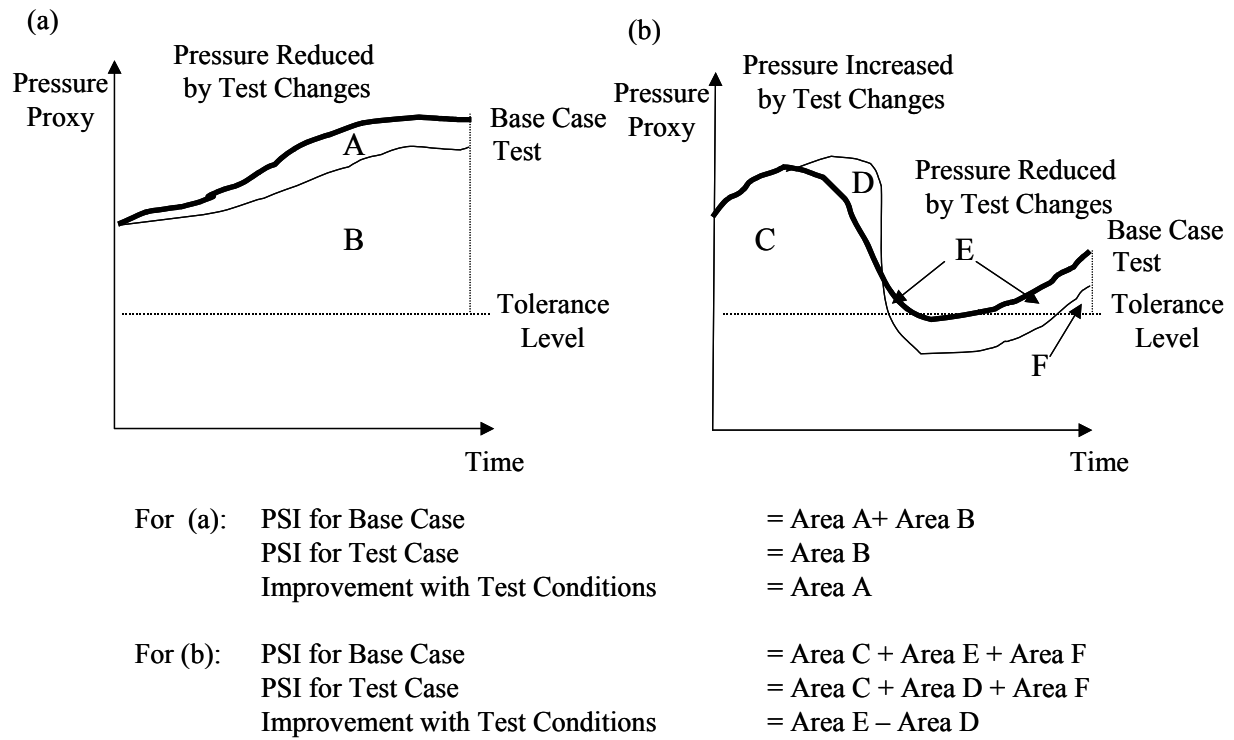


Figure 7

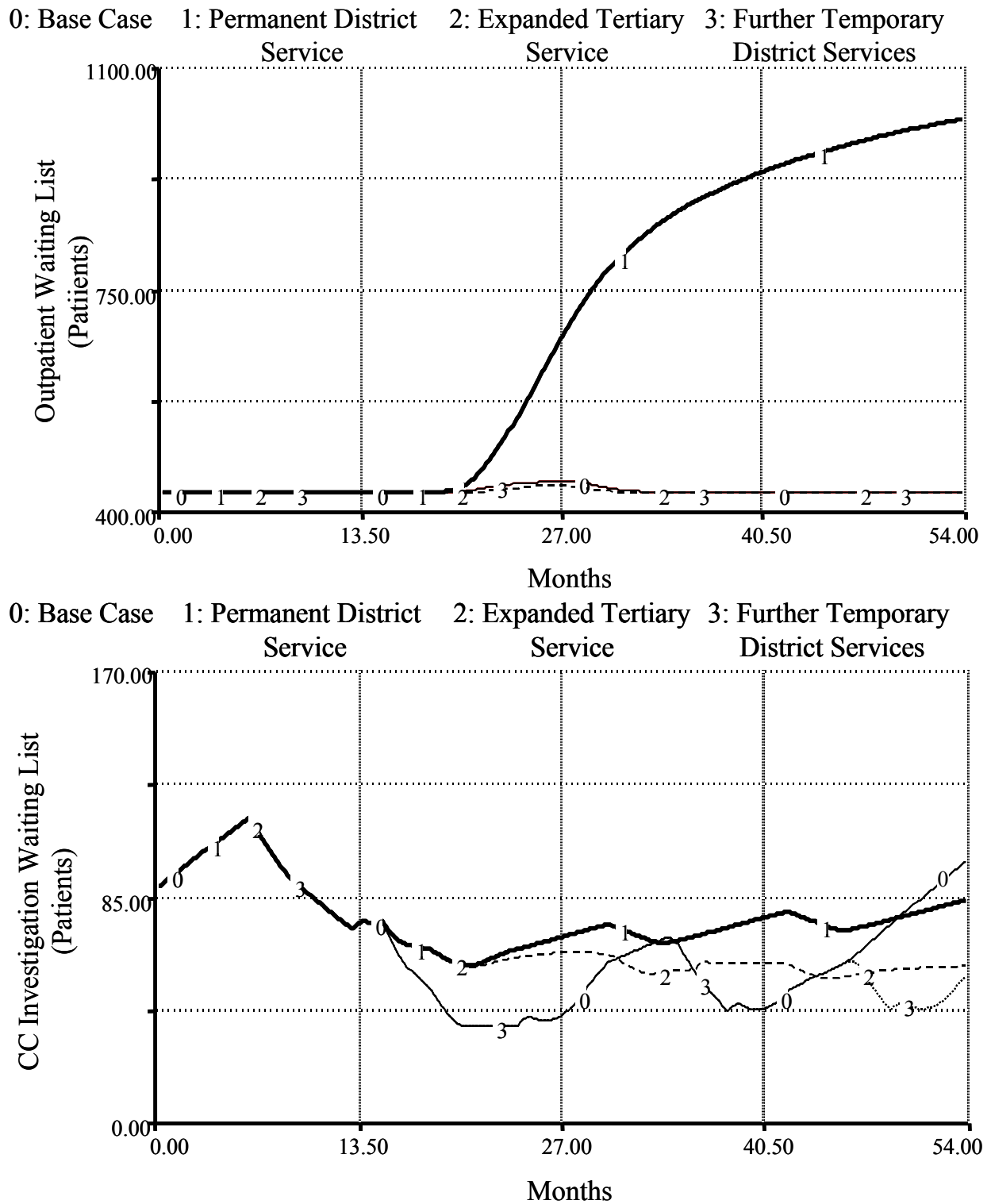
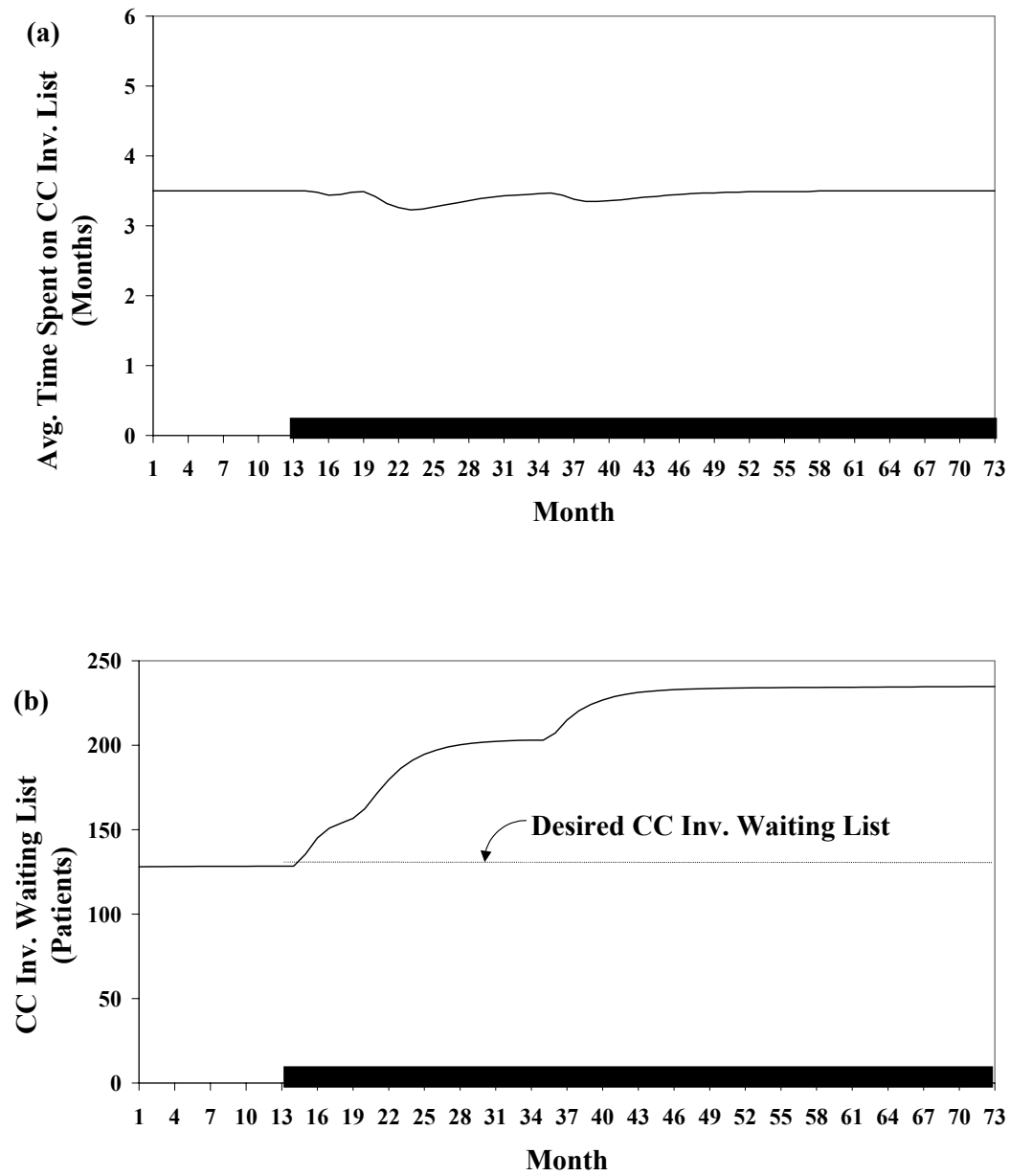
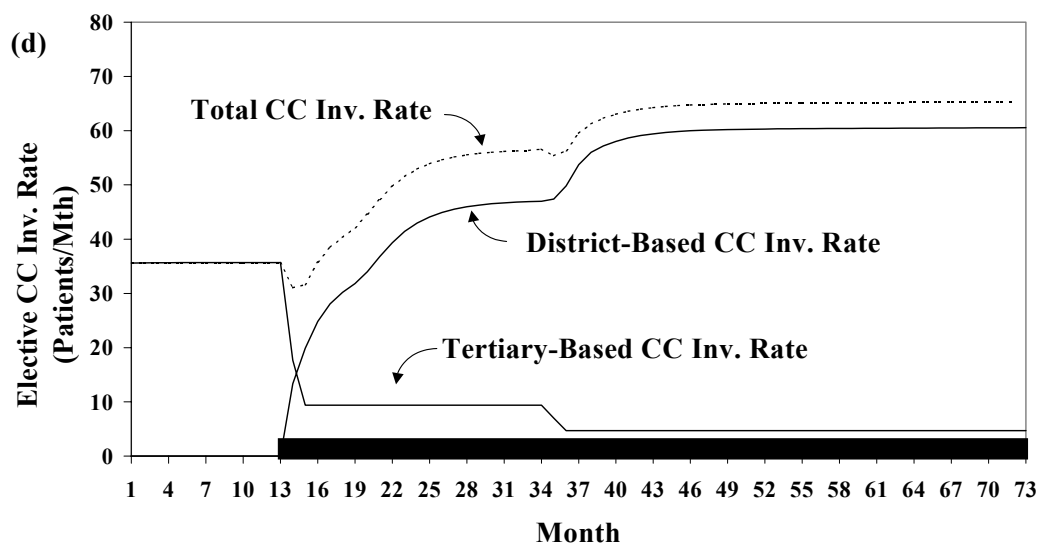
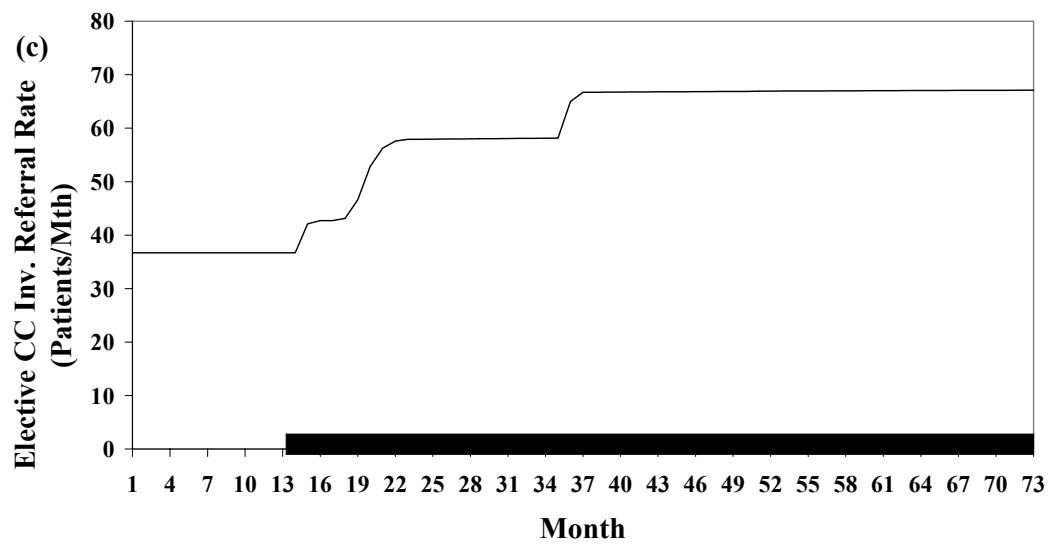


Figure 8

(Figure 9 has been enlarged and covers 3 pages)







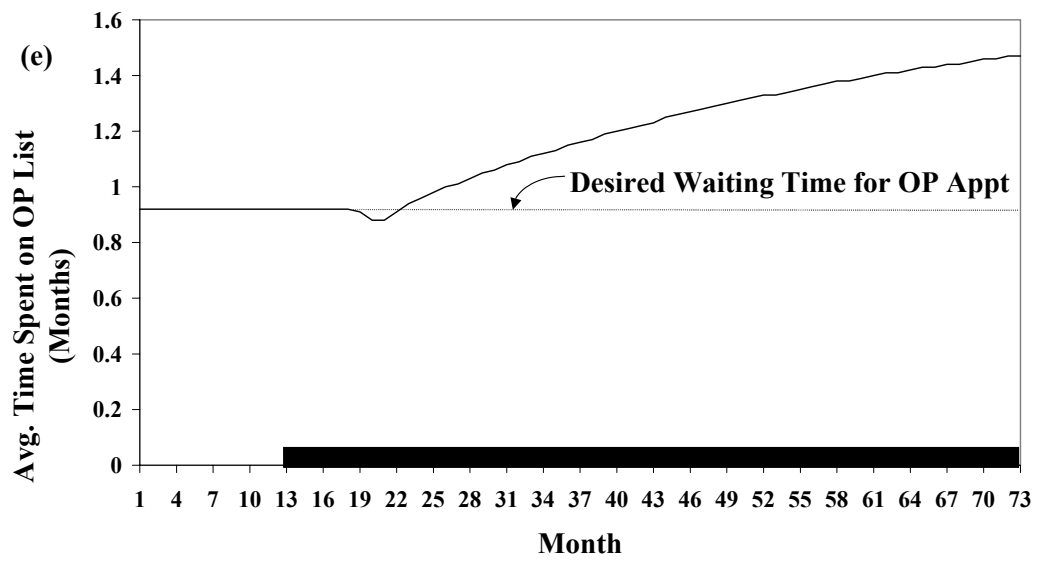


Figure 9

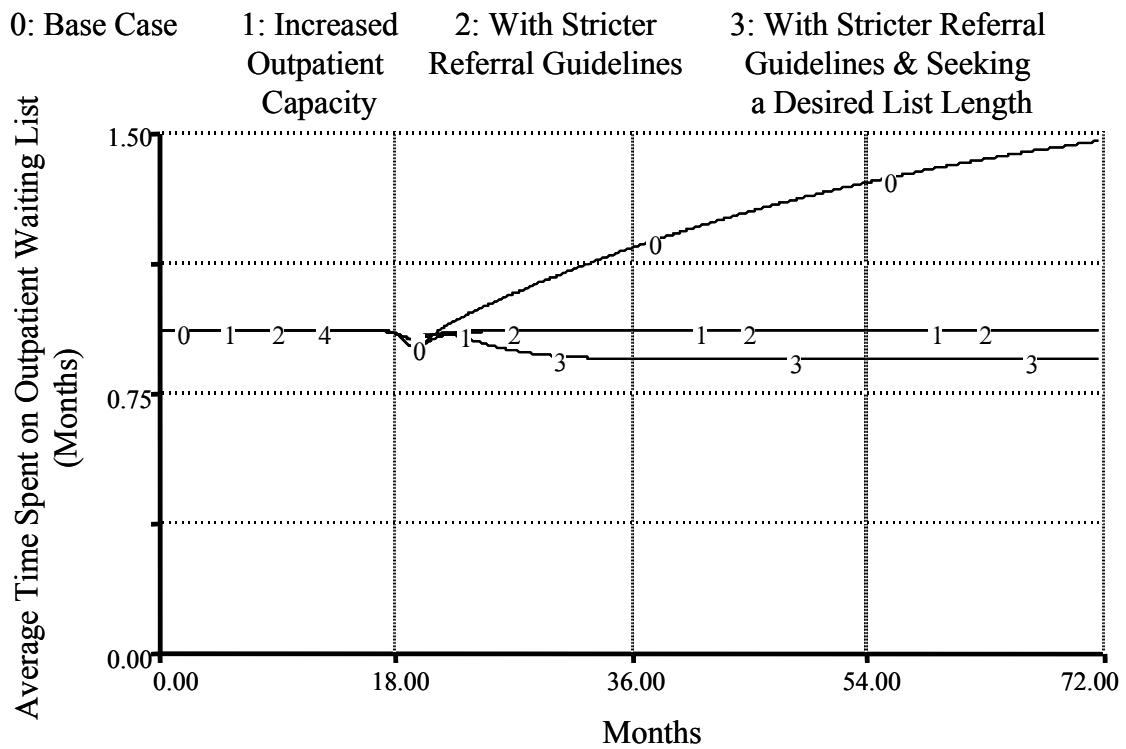
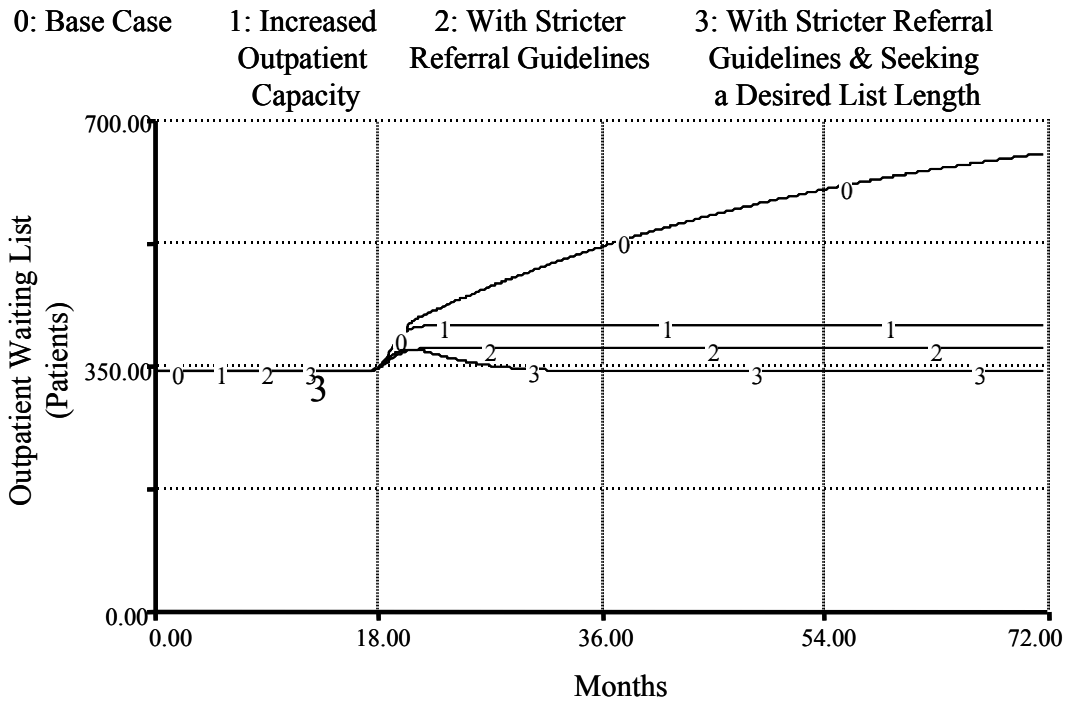


Figure 10

**Table 1**

Fig.	MAPE	Theil's Inequality Statistics*			R <sup>2</sup>	N
		Bias	Unequal Variation	Unequal Covariation		
5a	0.18	0.60	0.09	0.31	0.88	35
5b	0.13	0.68	0.01	0.31	0.93	35
5c	0.16	0.02	0.04	0.94	0.24	35
5d	0.16	0.04	0.09	0.87	0.80	35
5e	0.10	0.08	0.11	0.80	0.96	35
5f	0.30	0.00	0.03	0.97	0.91	35

**Table 2**

Performance Measure	Base Case	% Change from Base Case		
		Permanent District Service	Expanded Tertiary Service	Further Temporary District Services
PSI for outpatient waiting list	155	>+8000	-44	0
PSI for CC waiting list	891	+1.6	-40.5	-10.6
Cumul. referrals for an outpatient appt	19,937	+9.7	-0.2	0
Cumul. referrals for a CC investigation	947	+13.4	+4.7	-1.4
Cumul. outpatient activity	18,784	+3.5	-0.2	0
Cumul. CC investigations	848	+14.9	+9.4	+4.3
Cumul. Costs	2,439,461	+6.4	+1.7	+1

**Table 3**

Performance Measure	Base Case	% Change from Base Case		
		Increased Outpatient Capacity	With Stricter Referral Guidelines	With Stricter Referral Guidelines & Seeking a Desired List Length
PSI for outpatient waiting list	11,263	-68.4	-84.2	-98
PSI for time spent on outpatient waiting list	17	-100	-100	-100
PSI for CC inv. Waiting list	5,043	+2.4	-52.2	-96.6
Cumul. referrals for an outpatient appt	30,295	0	-6.5	-6.4
Cumul. referrals for a CC inv.	4,140	+0.9	-19.6	-19.5
Cumul. outpatient activity	29,092	+1.5	-5	-4.7
Cumul. CC investigations	3,922	+0.9	-18.6	-16.8
Cumul. Costs	6,801,121	+1	-10.2	-9.3

**Figure 1** Consequences of district services: improving access and stimulating demand. Black blocks indicate the use of district services

**Figure 2** The main patient flows. The direction of flow is indicated by the white arrowheads.

**Figure 3** The main sectors

**Figure 4** Base case waiting time, knowledge and skills referral multipliers \* - Neither stimulates nor suppresses demand; † - 100 represents a fully skilled CC operator; Values above 1 - Stimulation of demand; Values below 1 - Suppression of demand

**Figure 5** Basic feedback structure. B1 - Activity adjustment; B2 - Waiting time effect on demand; B3 - Other outpatient waiting list removals; R1 - Skills effect on demand; R2a&b - Knowledge effects on demand (for elective CC investigations and outpatient appointments respectively).

**Figure 6** Historical fit for the CC variables in the Ribsley case. Actual data is smoothed with a 2 point-centred moving average. Black blocks indicate the use of district services during month 14 to month 23 and month 34 to month 38; Avg - Average; CC Inv - CC Investigation

**Figure 7** Using a pressure summary index to measure an improvement under test conditions. PSI - Pressure summary index; Pressure Proxy - Waiting list length or average waiting time

**Figure 8** Increasing elective CC capacity for Ribsley case. The outpatient waiting list graphs for runs 0 and 3 are the same

**Figure 9** Selected variables for the Veinbridge base case scenario. Black blocks indicate the use of district services (introduced at month 13 and with integrated catheter laboratory opened at month 34); Avg - Average; CC Inv - CC Investigation; OP - Outpatient

**Figure 10** Meeting the outpatient waiting list and waiting time targets for Veinbridge case

**Table 1** Statistics of the historical fit for the CC variables in the Ribsley case. \* The failure to sum to unity is due to rounding errors.

**Table 2** Selected summary statistics of Ribsley policy runs. Cumul. - Cumulative; PSI - Pressure summary index

**Table 3** Selected summary statistics of Veinbridge policy runs. CC inv. - CC investigation; Cumul. - Cumulative; PSI - Pressure summary index.